



# PROJET DE FIN D'ANNEE

*Présenté à*

**La Faculté des Sciences de Sfax**

*Elaboré par*

**Akram HAGGUI**

**Malek LOGHMARI**

---

**Classification des Troubles Neuro-dégénératifs à Partir de Données  
de Marche**

---

*Soutenu le 29 mai 2025, devant le jury composé de :*

**M. Ali SALEM**

*Examineur*

**Mme. Hala BEZINE**

*Encadrant  
Académique*

Année Universitaire : 2024 – 2025

## Table des matières

1. Introduction .....	3
2. Analyse Comparative des Performances des Méthodes Existantes .....	4
3. Méthodologie .....	5
3.1. Description du Dataset .....	5
3.2. Prétraitement des données .....	7
3.3. Préparation des données pour l'entraînement .....	7
3.4. Modèles GAN et architecture proposée.....	9
3.4.1. Modèles GAN.....	9
3.4.2. Architecture proposée.....	9
3.5. Architecture Proposée pour la Classification .....	10
4. Résultats expérimentaux et discussion.....	12
4.1. Configuration expérimentale.....	12
4.2. Analyse comparative .....	13
3.4.3. Données générées vs données réelles .....	13
3.4.4. Mesures d'évaluation .....	14
4.3. Résultats de classification.....	17
5. Comparaison avec des travaux existants .....	18
6. Conclusion et perspectives .....	20

# 1. Introduction

Les maladies neurodégénératives sont des maladies qui endommagent lentement le cerveau et le système nerveux au fil du temps, ils affectent la manière dont les gens se déplacent et pensent. Trois exemples bien connus sont la sclérose latérale amyotrophique (ALS), Parkinson (PD) et Huntington (HD). Ces affections modifient souvent la façon dont une personne marche, ce que l'on appelle en anglais Gait. Dans ALS, les nerfs qui contrôlent les muscles cessent de fonctionner normalement, entraînant une faiblesse et une raideur musculaires. Cela rend la marche lente, instable et difficile : les gens peuvent traîner les pieds ou avoir du mal à les soulever. Dans la maladie de Parkinson, une partie du cerveau qui aide à contrôler le mouvement ne produit pas assez de dopamine, une substance chimique qui aide les muscles à fonctionner correctement. En conséquence, les gens peuvent marcher par petits pas, se déplacer lentement et parfois se figer sur place en essayant de marcher. La maladie de Huntington est une condition génétique qui cause des dommages au cerveau, entraînant des mouvements incontrôlés appelés chorée. Les personnes atteintes de la maladie de Huntington ont souvent un schéma de marche très irrégulier et imprévisible, avec des mouvements brusques, un mauvais équilibre et un risque accru de chutes. Même si ces maladies sont différentes, elles rendent toutes la marche plus difficile à leur manière et peuvent considérablement affecter l'indépendance et la vie quotidienne d'une personne.

Le tableau 1 présente une comparaison des caractéristiques de la marche associées aux trois maladies neurodégénératives.

*Tableau 1 : Comparaison des caractéristiques de la marche.*

Maladie neurodégénérative	Symptômes	Caractéristiques de la marche
Maladie de Parkinson (PD) [1]	Lenteur des mouvements, rigidité musculaire, tremblements, accélération involontaire des pas, difficulté à ajuster la posture pour éviter les chutes et blocage soudain des mouvements	Vitesse de marche réduite, cadence réduite, longueur de pas réduite, temps de swing réduit, temps de double support plus élevé
Sclérose Latérale Amyotrophique (ALS) [2]	Variations irrégulières des mouvements, faiblesse alternée des jambes, pieds ou chevilles, faiblesse incontrôlée	Vitesse de marche réduite, longueur de pas réduite, variabilité de la longueur de pas augmentée, temps de stride augmenté
Maladie de Huntington (HD) [3]	Mouvements incontrôlés, troubles psychiatriques et perte des capacités de pensée	Vitesse de marche réduite, longueur de pas/strides réduite, phase de stance/swing réduite, temps de support simple réduit

## 2. Analyse Comparative des Performances des Méthodes Existantes

Le tableau 2 contient les « accuracy » en % de différentes méthodes d'apprentissage automatique et profond pour la classification des maladies neurodégénératives à partir de données de marche, comparés pour quatre catégories.

Tableau 2 : Comparaison de l'accuracy entre les littératures existantes utilisant le même jeu de données.

Référence	Méthode	Accuracy			
		ALS vs Control	HD vs Control	PD vs Control	NDD vs Control
Bilgin et al. [4] (2017)	LDA, NB	90.93	–	–	–
Billgin et al. [5] (2018)	MLP, RBF, GR, SVM, DT	96.55	–	–	–
Elden et al. [6] (2018)	SVM	86.21	80.56	90.32	92.19
Paragliola et al. [7] (2021)	CNN, LSTM	85–90	75–82	80–95	80–86
Erda et al. [8] (2023)	CNN, LSTM	97.68	94.69	95.05	96.33
Joshi et al. [9] (2017)	SVM	–	–	90.32	–
Felix et al. [10] (2020)	SVM, KNN, DT	96.6	–	–	–
Felix et al. [11] (2019)	SVM, KNN, NB, LDA, DT	–	–	96.8	–
Ye et al. [12] (2018)	ANFIS	93.10	94.44	90.32	90.63
Begum et al. [13] (2020)	LDA, SVM, KNN, RF	90.63	90.63	90.63	92.19
Selzler et al. [14] (2018)	SVM, KNN	93.3	83.3	87.1	89.1
Faisal et al. [15] (2023)	NDDNet	96.67	97.22	96.30	98.2

### 3. Méthodologie

#### 3.1. Description du Dataset

Cette recherche repose sur la Base de Données 'Gait in Neurodegenerative Disease Database', partagée par Hausdorff et al [17], disponible sur PhysioNet [18]. L'ensemble de données comprend des enregistrements de données de marche de 64 sujets. Ce groupe a été réparti en quatre catégories distinctes, chacune correspondant à une maladie neurodégénérative spécifique. Le premier groupe est composé de participants sains, 16 personnes (4 femmes et 2 hommes). Le deuxième regroupe 20 personnes atteintes de la maladie de Huntington (HD), avec une majorité de femmes (14 femmes et 6 hommes). Le troisième groupe concerne la maladie de Parkinson (PD), incluant 15 participants (10 hommes et 5 femmes). Enfin, le dernier groupe est constitué de 13 personnes atteintes de sclérose latérale amyotrophique (ALS), (10 hommes et 3 femmes). La collecte des données s'est déroulée à la clinique de neurologie de l'hôpital général du Massachusetts. Les participants doivent marcher de façon autonome pendant 5 minutes dans un couloir droit de 77 mètres. Les signaux enregistrés, à une fréquence d'échantillonnage de 300 Hz, ont ensuite été analysés à l'aide de l'algorithme de segmentation des cycles de marche développé par Hausdorff et ses collègues, afin d'obtenir les temps de contact au sol pour chaque pas. Grâce aux capteurs placés sous les pieds, il a été possible d'extraire, pour chaque participant, un ensemble de 13 caractéristiques spatiotemporelles liées à la marche expliquée ci-dessous dans le tableau 3.

Lien du jeu de données : <https://physionet.org/content/gaitnidd/1.0.0/>

*Tableau 3 : Ensemble de 13 caractéristiques spatiotemporelles extraite de la marche.*

Caractéristique	Explication
Elapsed Time (sec)	Temps total enregistré depuis le début de l'acquisition de la marche.
Left Stride Interval (sec)	Durée séparant deux appuis consécutifs du pied gauche sur le sol.
Right Stride Interval (sec)	Durée séparant deux appuis consécutifs du pied droit sur le sol.
Left Swing Interval (sec)	Durée pendant laquelle le pied gauche est en l'air, c'est-à-dire non en contact avec le sol.
Right Swing Interval (sec)	Durée pendant laquelle le pied droit est en l'air.
Left Swing Interval (% of stride)	Proportion du cycle de marche durant laquelle le pied gauche est en l'air.
Right Swing Interval (% of stride)	Proportion du cycle de marche durant laquelle le pied droit est en l'air.
Left Stance Interval (sec)	Durée pendant laquelle le pied gauche reste en contact avec le sol.
Right Stance Interval (sec)	Durée pendant laquelle le pied droit reste en contact avec le sol.
Left Stance Interval (% of stride)	Proportion du cycle de marche où le pied gauche est en appui au sol.
Right Stance Interval (% of stride)	Proportion du cycle de marche où le pied droit est en appui au sol.
Double Support Interval (sec)	Durée pendant laquelle les deux pieds sont simultanément en contact avec le sol.
Double Support Interval (% of stride)	Proportion du cycle de marche durant laquelle les deux pieds sont au sol en même temps.

La figure 1 schématise un cycle de marche complet (gait cycle), décomposant les phases clés telles que le double appui, l'appui simple et la phase de balancement pour chaque pied, offrant ainsi une représentation claire des différentes étapes qui rythment la marche d'un individu.

La figure 2 présente les signaux temporels de marche enregistrés à partir des capteurs placés sous les pieds d'un sujet du dataset(Huntington 1), montrant les variations des forces exercées par le pied gauche (en bleu) et le pied droit (en rouge)

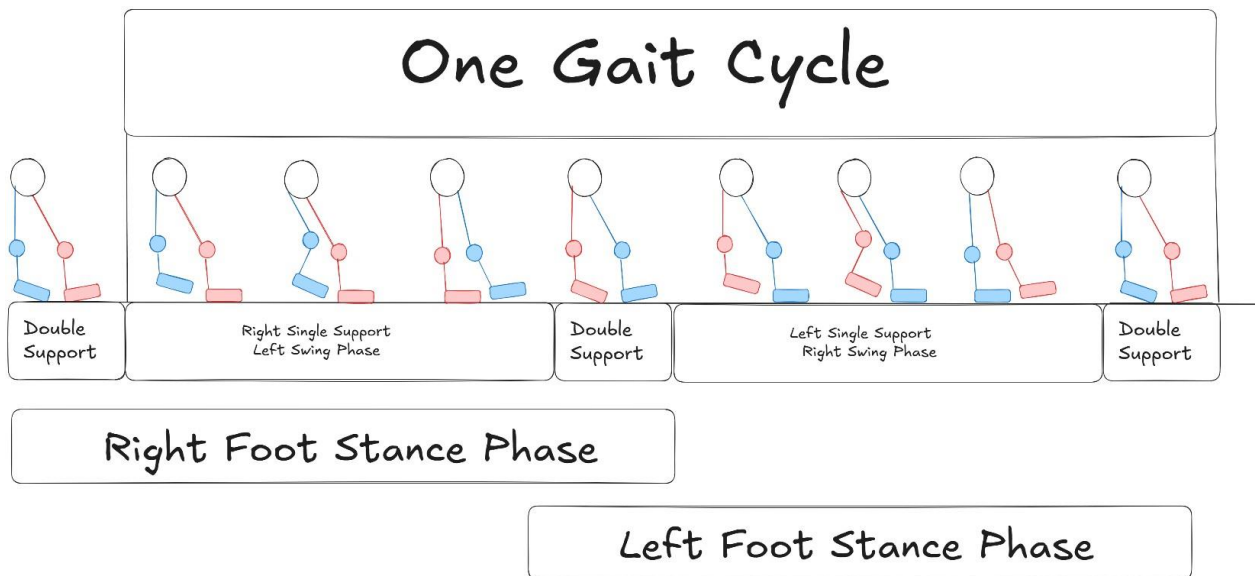


Figure 1 : Représentation schématique d'un cycle de marche complet.

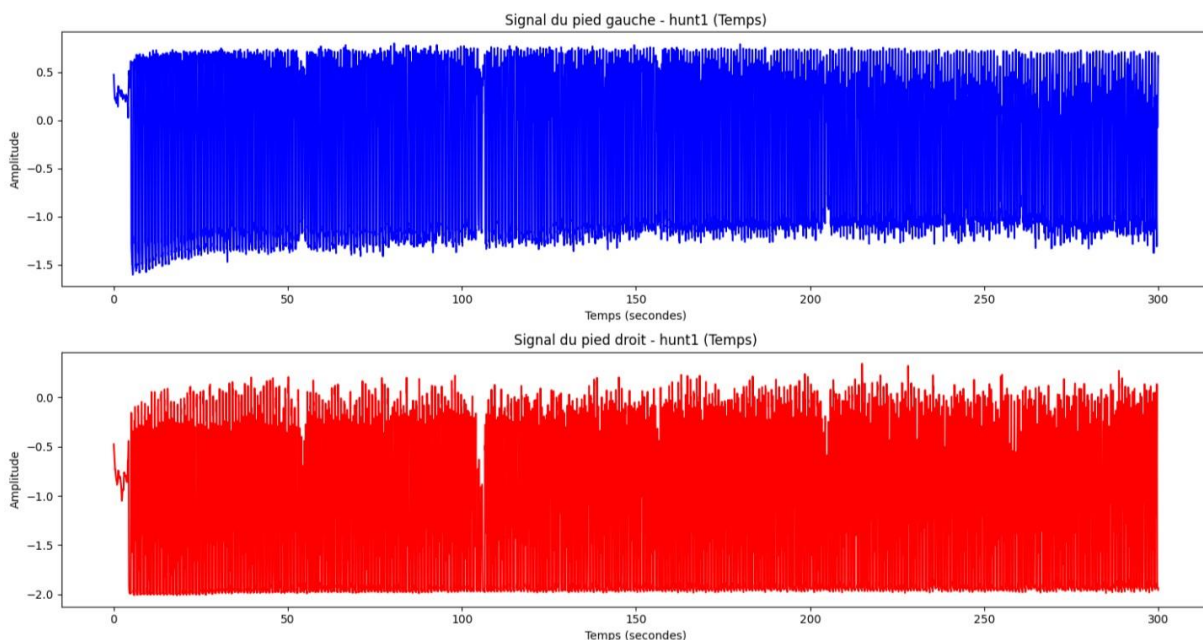


Figure 2 : Signaux temporels de marche pour un sujet de la maladie de Huntington.

### 3.2. Prétraitement des données

Avant d'exploiter les données, un prétraitement du signal s'est avéré nécessaire (Voir Figure 3). En effet, les enregistrements sont particulièrement sensibles au bruit ambiant et aux artefacts, ce qui peut introduire des informations parasites susceptibles de perturber les analyses ultérieures. Pour y remédier, un pipeline de prétraitement des données a été mis en place, incluant l'élimination des valeurs aberrantes, la filtration du bruit et la standardisation. La procédure de prétraitement est expliquée ci-dessous : Tout d'abord, les signaux de forces enregistrés tout au long de l'essai ont montré des incohérences attribuées à l'effet de démarrage. Pour résoudre ce problème, nous avons éliminé les incohérences en excluant les 20 premières secondes. Deuxièmement, un filtre passe-bas de Butterworth d'ordre deux avec une fréquence de coupure de 15 Hz a été appliqué au signal de force pour éliminer le bruit haute fréquence. Troisièmement, une normalisation min-max a été effectuée pour obtenir des amplitudes de signal standardisées entre  $[-1,1]$ . Enfin, un fenêtrage de 4 secondes avec 2 secondes de chevauchement a été appliqué. La contrainte de longueur du couloir a entraîné l'apparition de valeurs d'intervalle de foulée uniques ont été générés lorsque les sujets ont effectué un demi-tour. Pour aborder ce problème, la règle des 3 sigma a été utilisée lors du prétraitement, éliminant tous les échantillons qui dépassaient 3 écarts types ou moins de 3 écarts types par rapport à la valeur médiane globale.

### 3.3. Préparation des données pour l'entraînement

Après le prétraitement des signaux (filtrage, de normalisation, de découpage en fenêtres et d'élimination des valeurs aberrantes) une phase de structuration des données a été effectuée pour préparer les entrées destinées à l'entraînement du modèle GAN. Chaque fenêtre temporelle extraite (pied gauche ou du pied droit) a été associée à une étiquette correspondant à l'identifiant du sujet (Control, ALS, PD, HD). Cette étape d'étiquetage permet de faciliter l'organisation des classes pour l'entraînement et l'évaluation par le modèle génératif. Les données ont ensuite été redimensionnées pour adopter un format standardisé, compatible avec les architectures GAN spécialisées dans les séries temporelles. Chaque fenêtre a été convertie en un tuple de forme (1200, 1), représentant une séquence unidimensionnelle de 1200 points avec un seul canal. Cette structure tridimensionnelle (nombre d'échantillons, longueur de séquence, nombre de canaux) est essentielle pour assurer une bonne compatibilité avec notre modèle. L'ensemble de ce processus de préparation, incluant l'étiquetage et le redimensionnement des données, est illustré de manière schématique dans la Figure 3

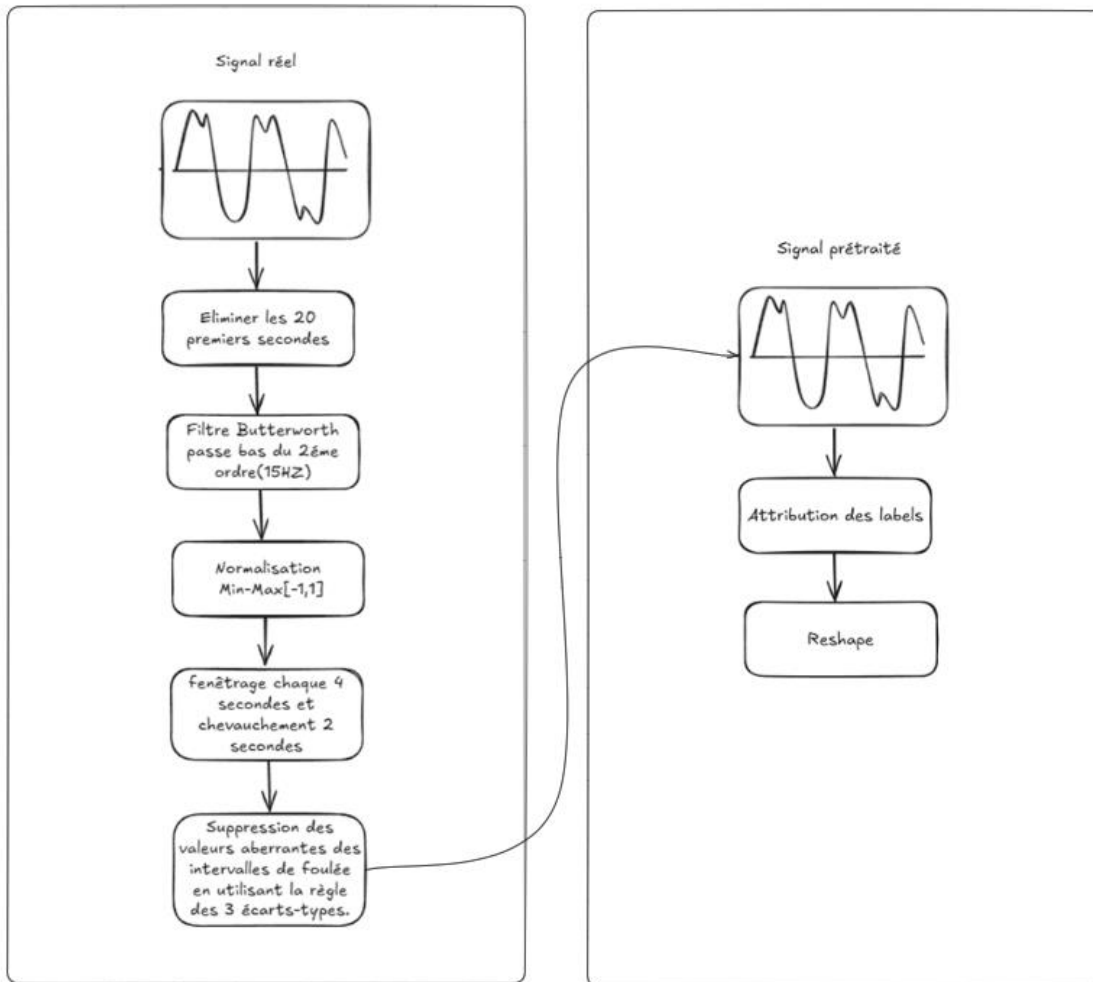


Figure 3 : Etapes de prétraitement et de préparation des signaux de marche.



### **3.4. Modèles GAN et architecture proposée**

#### **3.4.1. Modèles GAN**

Dans notre travail, on s'est appuyé sur les GAN (Generative Adversarial Networks), qui sont des modèles d'apprentissage profond composés de deux réseaux neuronaux qui « s'affrontent ». Le générateur essaie de produire des données qui ressemblent aux vraies, pendant que le discriminateur tente de repérer les fausses. Ce jeu de compétition permet aux deux réseaux de s'améliorer mutuellement, jusqu'à ce que les données créées soient presque impossibles à distinguer des vraies données.

Il existe une multitude de versions des GAN dont chacune est adaptée à un type de données ou à un objectif précis. Par exemple, les DCGAN sont souvent utilisés pour générer des images, les CGAN permettent de conditionner la génération à partir d'informations spécifiques, et les TCGAN, qui nous intéressent ici, sont spécialement conçus pour les données temporelles.

On a retenu cette dernière version (TCGAN) dans notre étude. Ce modèle nous aide à générer des signaux réalistes tout en respectant l'ordre et la structure dans le temps. Ce dernier nous permet d'enrichir notre jeu de données, notamment quand on a peu d'exemples ou un fort déséquilibre entre les différentes classes.

#### **3.4.2. Architecture proposée**

L'architecture se compose d'un générateur et d'un critic (Voir Figure 5). Le générateur va créer des signaux à partir de bruit. Ces derniers sont très proches des signaux réels. Pour le cas du critic ; qui correspond à un discriminateur dans un modèle GAN classique, il a le rôle d'attribuer un score (non une probabilité) aux signaux qu'il reçoit comme entrées (générées et réelles). Des scores élevés sont attribués aux vraies données et ceux faibles aux données générées. Le critic est entraîné pour maximiser la différence entre les scores des vraies et données faussées, dans le même temps le générateur apprend à produire des données faussées qui reçoivent des scores plus élevés.

Cette synergie entre le générateur et le critic continue jusqu'à ce que le générateur crée des données très proches des vraies.

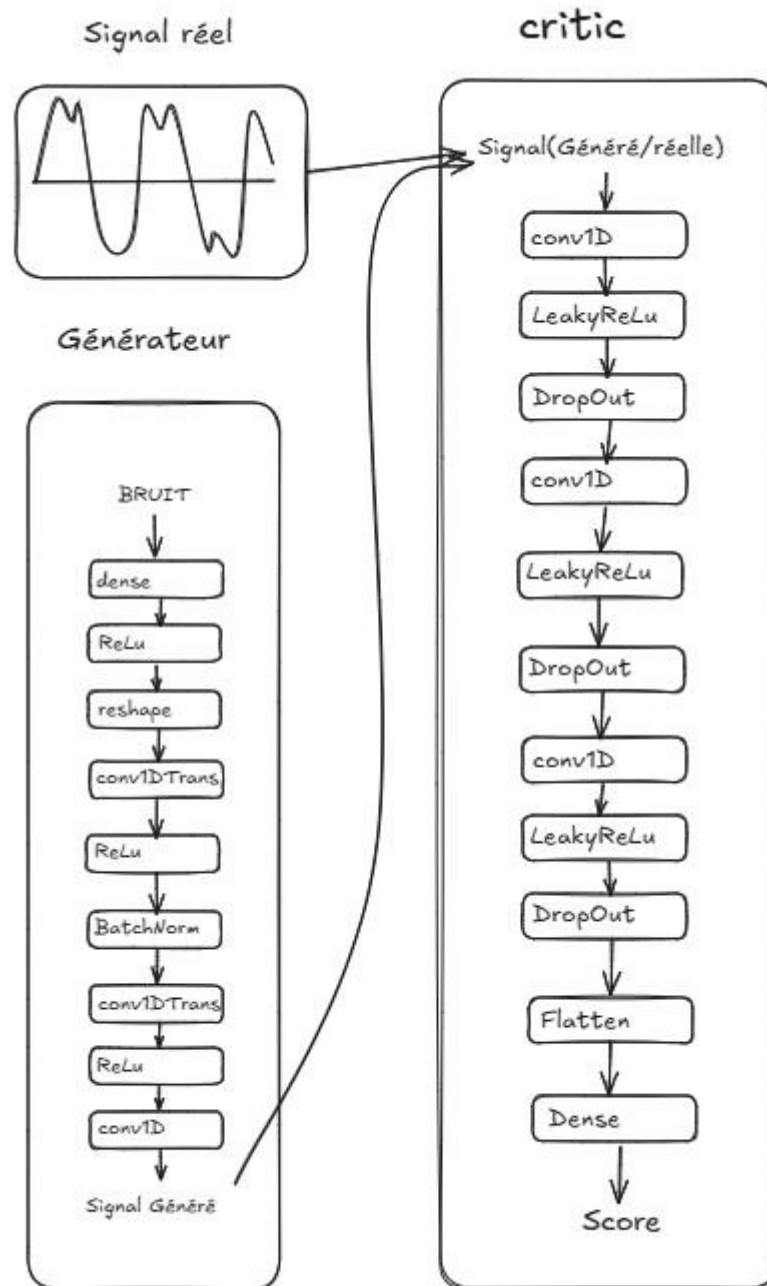


Figure 4 : Architecture du modèle GAN proposé.

### 3.5. Architecture Proposée pour la Classification

Dans notre approche, nous avons combiné deux types de réseaux (Voir Figure 6) : les réseaux de neurones convolutionnels (CNN) et les réseaux LSTM. L'idée est simple : le CNN repère automatiquement les motifs intéressants dans les signaux, un peu comme s'il balayait les données pour en extraire les formes ou variations importantes. Puis, c'est le LSTM qui prend

le relais. Son rôle est de suivre comment les motifs repérés évoluent avec le temps ; un point essentiel quand on travaille avec des signaux biologiques comme ceux de la marche. Grâce à la combinaison du CNN et du LSTM, notre modèle parvient à détecter des différences parfois très fines entre la démarche d'une personne en bonne santé et celle d'une personne atteinte d'un trouble neurologique, en analysant à la fois le contenu du signal et son évolution.

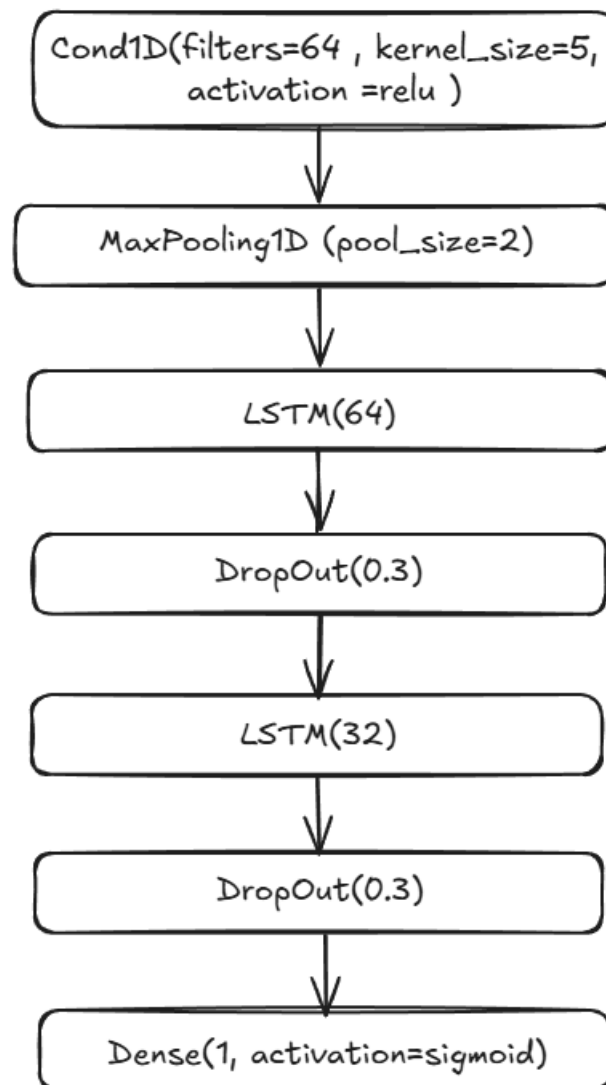


Figure 5 : Architecture du model CNN, LSTM.

## 4. Résultats expérimentaux et discussion

### 4.1. Configuration expérimentale

Le modèle WGAN est codé avec Python en utilisant les bibliothèques Keras et TensorFlow. L'entraînement a été réalisé sur Google Colab, avec ces ressources matérielles : GPU Tesla T4, 12gb RAM. L'entraînement du modèle dure environ 40 heures.

Comme mentionné précédemment, la base de données de PhysioNet (Gait in Neurodegenerative Disease) regroupe quatre catégories : trois types de troubles neurodégénératifs — la maladie de Parkinson, la maladie de Huntington et la sclérose latérale amyotrophique (ALS) — ainsi qu'un groupe de sujets sains servant de contrôle. Dans notre étude, nous avons proposé une nouvelle méthode d'augmentation de données basée sur les séries temporelles, afin de relever le défi complexe que représente la génération de signaux de marche pour des patients atteints de troubles neurodégénératifs. Pour y parvenir, nous avons structuré notre approche en plusieurs étapes. Dans un premier temps, nous avons choisi d'entraîner le modèle uniquement sur les signaux issus des sujets sains ("controls"), en combinant les données des deux pieds, gauche et droit. L'idée était de permettre au générateur d'apprendre la structure générale des signaux de marche normaux, afin de le guider progressivement vers la production de signaux réalistes. Cette phase d'apprentissage constitue une base solide avant d'aborder des données plus spécifiques et variées liées aux pathologies.

Pour aller plus loin, nous avons divisé notre approche en plusieurs sous-problèmes en appliquant le transfer learning avec fine-tuning. Le modèle a ainsi été progressivement adapté aux différents groupes : sujets sains (controls), ALS, Huntington et Parkinson, afin qu'il puisse apprendre les particularités propres à chaque profil. Au départ, nous avons essayé de réaliser le fine-tuning en utilisant les données des deux pieds ensemble, dans l'espoir d'enrichir l'apprentissage du modèle. Mais, nous avons constaté que les résultats n'étaient pas aussi satisfaisants que prévu. C'est ce constat qui nous a amenés à changer de stratégie : nous avons décidé d'entraîner le modèle séparément pour chaque pied, en traitant indépendamment les données du pied gauche et du pied droit, et ça pour chaque sujet — qu'il soit sain ou atteint d'un trouble.

Toutes les informations détaillées concernant les hyperparamètres utilisés lors de l'entraînement des modèles sont présentées et synthétisées dans le tableau 4.

*Tableau 4 : Tableau des hyperparamètres d'entraînement.*

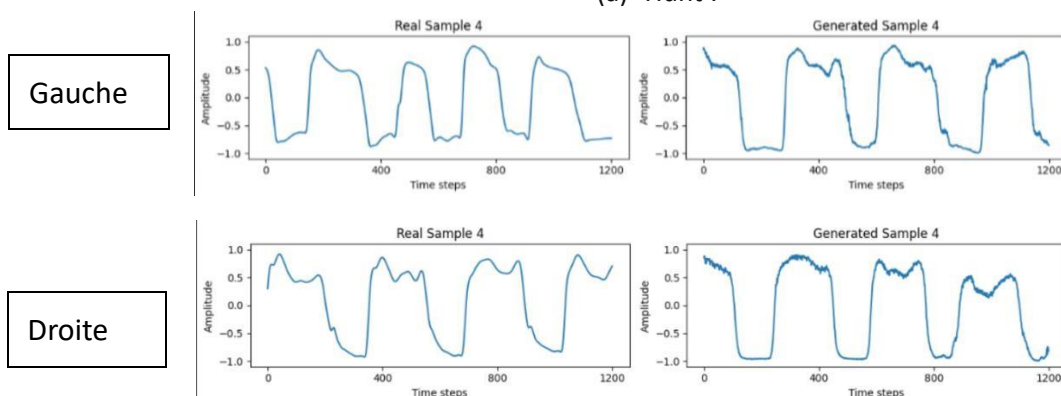
Hyperparamètres	Valeur
Learning rate	$2 \times 10^{-4}$
Learning Algorithm	Adaptive moment estimation (Adam)
Batch Size	64
Epochs	300
DropOut	0.25
Regularization	Gradient Penalty

## 4.2. Analyse comparative

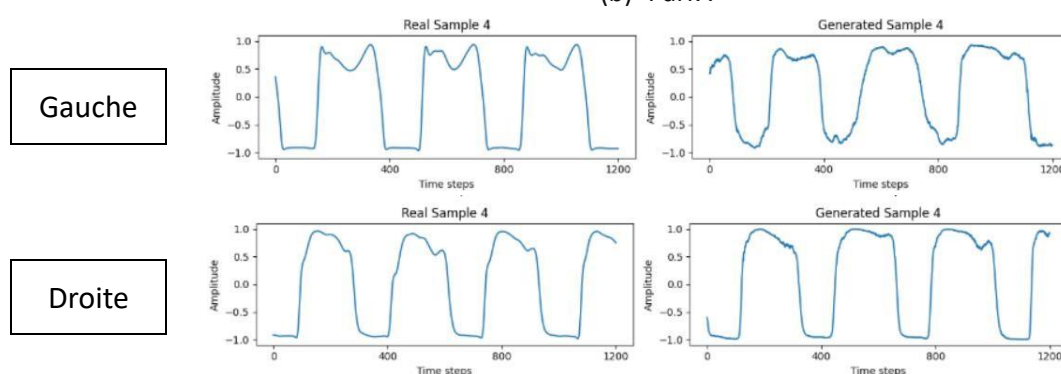
### 3.4.3. Données générées vs données réelles

Comme illustré dans la Figure 7, nous présentons des visualisations comparatives entre des fenêtres de données réelles et des fenêtres de données générées, pour les quatre groupes de sujets (Huntington, Parkinson, ALS et contrôles), et ce, séparément pour le pied gauche et le pied droit.

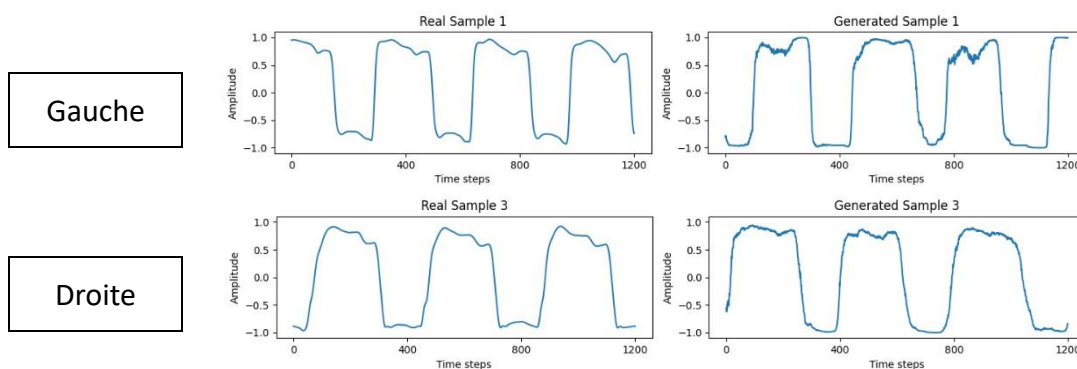
(a) Hunt :



(b) Park :



(c) Als :



(d) Control :

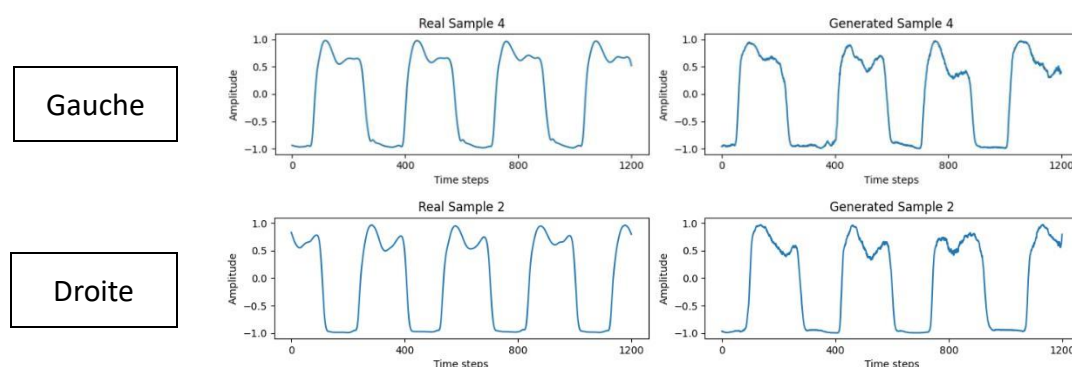


Figure 6 : Comparaisons visuelles des signaux réels vs générés (Hunt, Park, ALS, Control).

D'après les résultats, on voit que le GAN arrive plutôt bien à reproduire la forme générale des signaux de marche, en suivant les variations principales. Mais on peut aussi constater que les signaux générés restent assez différents des vrais, ce qui montre que le modèle n'a pas encore totalement réussi à imiter la distribution réelle des données. Il y a donc encore une bonne marge pour améliorer la qualité des signaux produits

#### 3.4.4. Mesures d'évaluation

Les résultats du Tableau 5 nous montrent que le modèle arrive plutôt bien à imiter les mouvements des personnes sans troubles (les contrôles), surtout du côté gauche. Les signaux qu'il génère ressemblent beaucoup aux vrais, que ce soit en forme ou en structure. En revanche, dès qu'on passe aux patients atteints de maladies comme la ALS, Huntington (HUNT) ou Parkinson (PARK), les choses se compliquent. Même si la distribution générale des signaux reste proche, le modèle a plus de mal à bien reproduire les détails dans le temps : les signaux générés ne sont pas toujours bien alignés avec les vrais, et ils ne suivent pas toujours la même forme. Cela se voit notamment avec des distances DTW plus grandes et des corrélations parfois négatives. En clair, le modèle sait bien imiter ce qu'il connaît bien (les sujets sains), mais il peine à capter les particularités des mouvements pathologiques, surtout dans les cas les plus graves comme la ALS.

Tableau 5 : Tableau des métriques d'évaluation.

	MMD	DTW	Cross-Corr
Control (gauche)	0.0200	9.8999	0.0867
Control (droite)	0.0201	10.6667	0.0085
ALS (gauche)	0.0202	12.8681	-0.0028
ALS (droite)	0.0212	13.6184	-0.0246
HUNT (gauche)	0.0200	10.8602	0.0296
HUNT (droite)	0.0200	11.9223	-0.1036
PARK (gauche)	0.0200	11.8066	-0.0146
PARK (droite)	0.0200	10.9519	0.0345

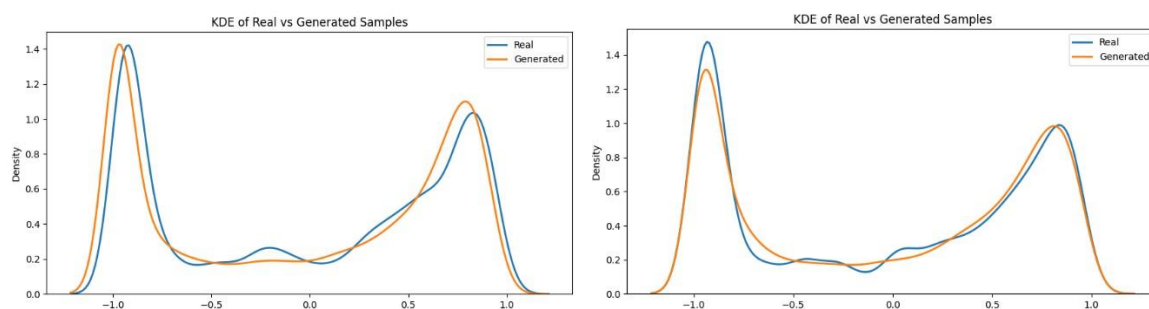
MMD (Maximum Mean Discrepancy) : La discrédance moyenne maximale est une mesure statistique qui compare la distribution de deux échantillons de données en utilisant une classe de fonctions appelées noyaux. Elle est souvent utilisée comme un test d'hypothèse non paramétrique pour déterminer si deux échantillons proviennent de la même distribution ou non.

DTW (Dynamic Time Warping): La déformation temporelle dynamique est un algorithme permettant de mesurer la similarité entre deux suites qui peuvent varier au cours du temps.

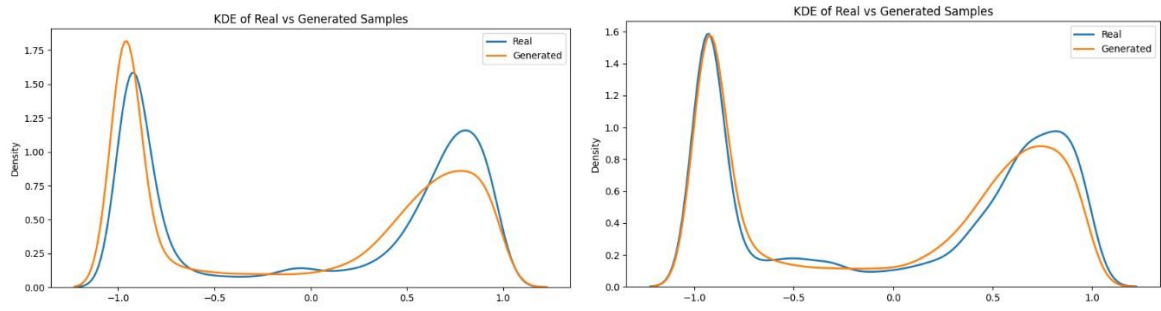
Cross-Correlation Moyenne: La corrélation croisée est une technique statistique utilisée pour mesurer la relation entre deux ensembles de données de séries chronologiques. Elle quantifie le degré de corrélation entre une série et une autre série à différents décalages temporels

Figure 7 : Estimation de la densité (KDE) des échantillons réels vs générés.

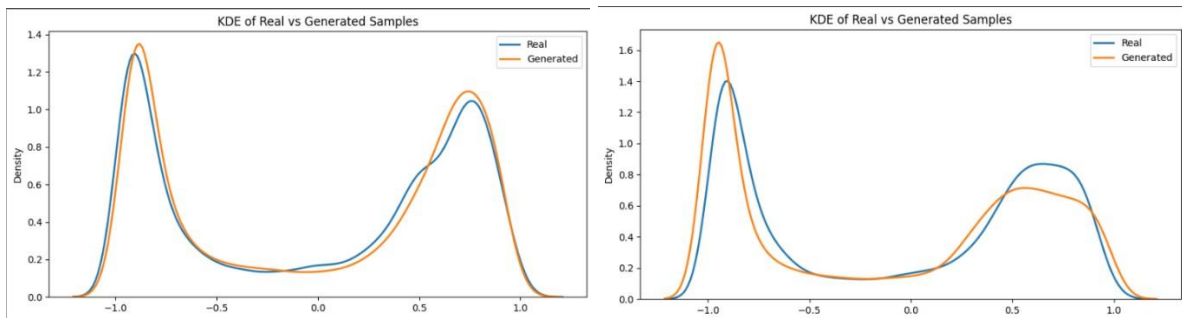
(a) ALS



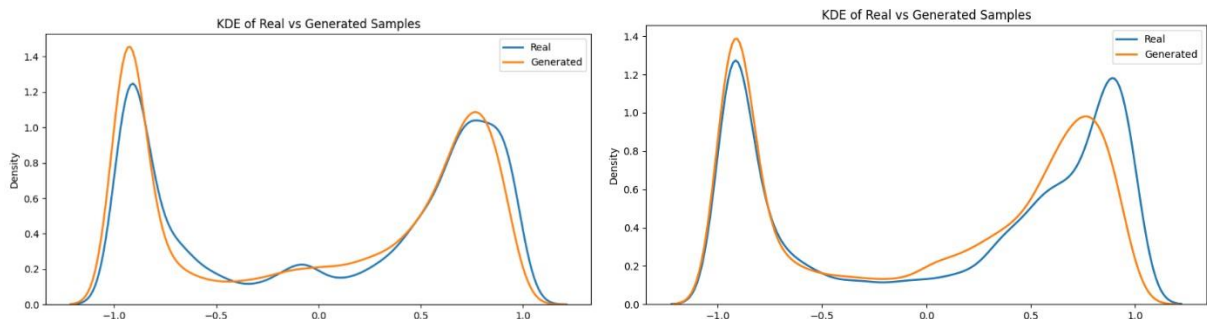
(b) Control



(c) Hunt



(d) Park



Les graphiques ci-dessus Figure 8 comparent les distributions des signaux réels (en bleu) et générés (en orange) à l'aide de courbes KDE, pour quatre cas différents : (a) ALS, (b) sujets sains (Control), (c) Huntington et (d) Parkinson. Grâce à ces visualisations, on peut se faire une idée assez claire de la capacité du modèle GAN à imiter les signaux réels. Dans le cas des sujets sains (b) Control, les courbes sont très proches l'une de l'autre, ce qui suggère que le modèle a réussi à bien reproduire la forme générale et les variations typiques de ces signaux. Pour les cas ALS et Parkinson, on observe quelques petites différences par endroits, mais dans l'ensemble, les résultats restent cohérents et plutôt satisfaisants. En revanche, pour Huntington, les différences sont plus visibles : les courbes ne se recoupent pas aussi bien, ce qui suggère que le modèle a eu plus de mal à capter la structure de ces signaux. Ces résultats suggèrent que le modèle s'en sort assez bien pour certains groupes, notamment les sujets



sains, mais qu'il a encore du mal à bien reproduire les signaux de cas plus complexes, comme ceux liés à la maladie de Huntington. Cela montre qu'il y a encore une marge d'amélioration possible.

### 4.3. Résultats de classification

Avant d'évaluer les performances de notre modèle, nous avons pris soin de rééquilibrer notre base de données. À l'origine, certaines catégories étaient sous-représentées — par exemple, on ne comptait que 13 enregistrements pour la maladie de ALS, 15 pour Parkinson, et 16 pour les sujets Control, contre 20 pour le groupe Hunt. Pour pallier ce déséquilibre, nous avons fait appel à notre modèle GAN, qui nous a permis de générer des données supplémentaires de manière réaliste dans les classes les moins fournies. Grâce à cette démarche, nous avons obtenu une base plus équilibrée : 18 enregistrements pour Parkinson, 20 pour Hunt, 16 pour ALS et 19 pour Control. Ce nouvel équilibre nous permet d'évaluer notre modèle de manière plus juste et représentative.

Tableau 6 : Résultats de classification obtenus pour différents NDD(ALS,Park,Hunt).

Modele	Accuracy	Sujet	Precision	Recall	F1-score
Control vs ALS	0.97	Control	0.96	0.98	0.97
		ALS	0.97	0.95	0.96
Control vs Park	0.94	Control	0.93	0.95	0.94
		Park	0.95	0.92	0.94
Control vs Hunt	0.96	Control	0.96	0.96	0.96
		Hunt	0.96	0.95	0.96
Control vs NDD	0.97	Control	0.93	0.93	0.93
		NDD	0.98	0.98	0.98

En partant d'une base de données équilibrée grâce au modèle GAN, nos modèles montrent une très bonne capacité à différencier les personnes en bonne santé de celles souffrant de maladies neurodégénératives, avec une précision qui peut atteindre 97 % (Voir Tableau 6). Les différentes mesures (précision, rappel, F1-score) confirment que ces modèles sont solides, que ce soit pour détecter la sclérose latérale amyotrophique (ALS), la maladie de Parkinson (Park) ou la maladie de Huntington (Hunt). Parmi eux, le modèle qui sépare le groupe Control des NDD est particulièrement performant, avec un F1-score de 0,98, ce qui montre qu'il est fiable. Ces résultats sont prometteurs et montrent que ce type d'outil pourrait vraiment aider les médecins en fournissant un diagnostic précis et cohérent.

## **5. Comparaison avec des travaux existants**

Un travail qui se rapproche beaucoup de notre projet a été réalisé par Albert et al. [19], qui ont proposé une méthode d'augmentation de données kinématiques à l'aide de réseaux antagonistes génératifs (GANs). Leur objectif était d'améliorer la classification de patients atteints de maladies neurologiques, comme l'AVC et la maladie de Parkinson, en générant des mouvements réalistes (principalement des squats) à partir de données existantes. Pour cela, ils ont utilisé une architecture de type HP-GAN combinée à des couches GRU, et ont entraîné leur modèle sur le dataset KIMORE, composé de données enregistrées en laboratoire. Les résultats montrent que l'ajout de données synthétiques a permis d'augmenter la précision du modèle de classification.

De notre côté, nous avons suivi une démarche similaire, mais avec plusieurs différences importantes. Tout d'abord, nous avons utilisé un Wasserstein GAN (WGAN) adapté aux séries temporelles, ce qui est plus approprié pour le type de données que nous avons utilisées, à savoir des signaux de marche (gait). Contrairement à leur travail qui se concentre sur des exercices spécifiques réalisés dans un cadre contrôlé, notre projet s'appuie sur des données enregistrées lors de marches naturelles, à l'aide de capteurs inertiels. Ensuite, au lieu de nous limiter à deux pathologies, nous avons travaillé sur un ensemble plus large de troubles neurologiques, en incluant des sujets atteints de Parkinson, de ALS, de la maladie de Huntington, ainsi qu'un groupe contrôle. Enfin, nous avons utilisé un dataset différent, le Gait in Neurodegenerative Disease Database (disponible sur PhysioNet), qui contient des enregistrements plus variés et plus représentatifs de la marche dans un contexte réel.

Toutes ces différences rendent la comparaison intéressante, car elles montrent comment les GANs peuvent être adaptés à des contextes variés, et soulèvent aussi la question de leur efficacité selon le type de données et les pathologies étudiées.

Les résultats obtenus dans notre étude présentent des performances globales très compétitives par rapport aux travaux antérieurs (Tableau 7). Par exemple, pour la classification

entre les sujets sains et les patients atteints de ALS, notre modèle atteint une précision de 0.97, ce qui surpasse la plupart des méthodes présentées dans la littérature, notamment celles de Bilgin et al. (2018) avec 96.55 % et même celles de Felix et al. (2020) avec 96.6 %. De même, pour la détection de la maladie de Parkinson (PD), notre précision atteint 0.94, proche voire équivalente aux meilleurs scores rapportés par Erda et al. (2023) (95.05 %) et Felix et al. (2019) (96.8 %), tout en conservant une grande stabilité entre les classes. Concernant la maladie de Huntington (HD), nous obtenons une précision constante de 0.96, comparable aux performances les plus élevées recensées, comme celles de Ye et al. (2018) avec 94.44 %. Enfin, pour la classification entre sujets sains et patients atteints de troubles neurodégénératifs généraux (NDD), notre modèle atteint 0.97, rivalisant avec le système NDDNet de Faisal et al. (2023) qui affiche 98.2 %. Ces résultats mettent en évidence la robustesse et la fiabilité de notre approche, tout en soulignant sa capacité à généraliser efficacement sur différentes pathologies neurologiques.

Tableau 7 : Comparaison de l'accuracy entre les littératures existantes.

Référence	Méthode	Accuracy			
		ALS vs Control	HD vs Control	PD vs Control	NDD vs Control
Billgin et al. [5] (2018)	MLP, RBF, GR, SVM, DT	96.55	–	–	–
Paragliola et al. [7] (2021)	CNN, LSTM	85–90	75–82	80–95	80–86
Erda et al. [8] (2023)	CNN, LSTM	97.68	94.69	95.05	96.33
Felix et al. [10] (2020)	SVM, KNN, DT	96.6	–	–	–
Felix et al. [11] (2019)	SVM, KNN, NB, LDA, DT	–	–	96.8	–
Ye et al. [12] (2018)	ANFIS	93.10	94.44	90.32	90.63
Faisal et al. [15] (2023)	NDDNet	96.67	97.22	96.30	98.2
<b>Travail proposé</b>	<b>CNN, LSTM</b>	<b>97.04</b>	<b>96.85</b>	<b>94.47</b>	<b>97.24</b>

## 6. Conclusion et perspectives

Même si le modèle GAN n'atteint pas une perfection absolue, il a tout de même démontré une capacité précieuse à générer de nouveaux signaux réalistes. Ces signaux synthétiques ont permis d'augmenter significativement la base de données existante sur la marche chez les patients atteints de maladies neurodégénératives. En enrichissant ainsi notre jeu de données, nous avons pu entraîner les modèles de classification sur un ensemble plus équilibré et diversifié. Cette amélioration de la qualité et de la quantité des données a conduit à des performances de classification bien plus fiables et robustes, comme en témoignent les mesures d'évaluation obtenues. Ces résultats encouragent à poursuivre le développement et l'optimisation de ces techniques pour mieux soutenir la recherche et le diagnostic dans ce domaine.

Par ailleurs, pour aller plus loin, il serait intéressant d'explorer l'utilisation d'architectures encodeur-décodeur comme méthode alternative d'augmentation de données. Ces modèles pourraient permettre de mieux capturer les caractéristiques complexes des signaux et générer des données synthétiques encore plus précises et spécifiques. Cette approche pourrait enrichir davantage la base de données et renforcer la robustesse des modèles de classification, ouvrant la voie à des applications cliniques plus fiables.

## Références

1. A. Zanardi, E. da Silva, and R. C. et al., "Gait parameters of Parkinson's disease compared with healthy controls: a systematic review and meta analysis," *Scientific Reports*, vol. 11, 2021.
2. J. M. Hausdorff, A. Lertratanakul, M. E. Cudkowicz, A. L. Peterson, D. Kaliton, and A. L. Goldberger, "Dynamic markers of altered gait rhythm in Amyotrophic Lateral Sclerosis," *Journal of Applied Physiology*, vol. 88, no. 6, pp. 2045–2053, 2000.
3. S. J. Pyo, H. Kim, I. S. Kim, Y. M. Park, M. J. Kim, H. M. Lee, and S. B. Koh1, "Quantitative Gait Analysis in Patients with Huntington's Disease," *Journal of Movement Disorders*, vol. 10, no. 3, pp. 140–144, Sep. 2017.
4. Bilgin S (2017) The impact of feature extraction for the classification of amyotrophic lateral sclerosis among neurodegenerative diseases and healthy subjects. *Biomed Signal Process Control* 31:288–294
5. Bilgin S, Akin ZE, (2018) Gait pattern discrimination of ALS patients using classification methods. *Turk J Electr Eng Comput Sci* 26(3):1367–1377
6. Elden RH, Al-Atabany W, Ghoneim VF (2018) Gait variability analysis in neurodegenerative diseases using nonlinear dynamical modelling. In: 2018 9<sup>th</sup> Cairo international biomedical engineering conference (CIBEC). IEEE, pp 41–44
7. Paragliola G, Coronato A (2021) A deep learning-based approach for the classification of gait dynamics in subjects with a neurodegenerative disease, in: *Intelligent Systems*

and Applications: Proceedings of the 2020 Intelligent Systems Conference (Intel liSys) Volume 3. Springer, pp 452–468

8. Erdas ÇB, Sümer E, Kibaroglu S (2023) Neurodegenerative diseases detection and grading using gait dynamics. *Multimed Tools Appl* 1–18
9. Joshi D, Khajuria A, Joshi P (2017) An automatic non-invasive method for Parkinson's disease classification. *Comput Methods Prog Biomed* 145:135–145
10. Felix JP, do Nascimento HAD, Guimarães NN, Pires EDO, da Silva Vieira G, de Souza Alencar W (2020) An effective and automatic method to aid the diagnosis of amyotrophic lateral sclerosis using one minute of gait signal. In: 2020 IEEE International conference on bioinformatics and biomedicine (BIBM). IEEE, pp 2745–2751
11. FélixJP,Vieira FH,CardosoÁA,FerreiraMV,FrancoRA,Ribeiro MA, Araújo SG, Corrêa HP, Carneiro ML (2019) A Parkinson's disease classification method:an approach using gait dynamics and detrended fluctuation analysis.In:2019IEEECanadianconference of electrical and computer engineering (CCECE). IEEE, pp 1–4
12. Ye Q, Xia Y, Yao Z (2018) Classification of gait patterns in patients with neurodegenerative disease using adaptive neuro-fuzzy inference system. *Computational and mathematical methods in medicine* 2018
13. BegumSV,RaniMP(2020) Recognition of neurodegenerative diseases with gait patterns using double feature extraction methods. In: 2020 4th international conference on intelligent computing and control systems (ICICCS). IEEE, pp 332–338
14. Selzler R, Green JR, GoubranR(2018)Neurodegenerative disease prediction based on gait analysis signals acquired with force sensitive resistors. In: 2018 IEEE life sciences conference (LSC). IEEE, pp 122–125
15. Faisal MAA, Chowdhury ME, Mahbub ZB, Pedersen S, Ahmed MU, Khandakar A, Alhatou M, Nabil M, Ara I, Bhuiyan EH et al (2023) NDDNet:a deep learning model for predicting neurodegenerative diseases from gait pattern. *Appl Intell* 1–13
16. Fourati, J., Othmani, M., Salah, K.B. *et al.* A new parallel-path ConvMixer neural network for predicting neurodegenerative diseases from gait analysis. *Med Biol Eng Comput* (2025). <https://doi.org/10.1007/s11517-025-03334-w>
17. Hausdorff JM, Lertratanakul A, Cudkowicz ME, Peterson AL, Kaliton D, Goldberger AL. Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis. *J Applied Physiology*; 88:2045-2053, 2000.
18. Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., ... & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation* [Online]. 101 (23), pp. e215–e220.
19. J. Albert, P. Glöckner, B. Pfitzner, and B. Arnrich, "Data augmentation of kinematic time-series from rehabilitation exercises using GANs," \*2021 IEEE International Conference on Omni-Layer Intelligent Systems (COINS)\*, Barcelona, Spain, 2021, pp. 1–6, doi: 10.1109/COINS51742.2021.9524284.